An Information Fusion Framework for Context-based Accidents Prevention

Nayat Sánchez-Pi, Luis Martí, José Manuel Molina and Ana Cristina Bicharra García

***Abstract*—The oil and gas industry is increasingly concerned with achieving and demonstrating good performance with regard occupational health and safety (OHS) issues, through the control of its OHS risks, which is consistent with its core policy and objectives. There are standards to identify and record workplace accidents and incidents to provide guiding means on preven-tion efforts, indicating specific failures or reference, means of correction of conditions or circumstances that culminated in an accident. Therefore, events recognition is central to OHS, since the system can selectively start proper prediction services according to the user current situation and past knowledge taken from huge databases. In this sense, a fusion framework that combines data from multiples sources to achieve more specific inferences is needed. In this paper we propose a machine learning algorithm to learn from past anomalous events related to accident events in time and space. It also uses additional knowledge, like the contextual knowledge: user profile, event location and time, etc. Our proposed model provides the big picture about risk analysis for that employee at that place in that moment in a real world environment. Our main contribution lies in building a causality model for accident investigation by means of well-defined spatiotemporal constraints in the offshore oil industry domain.**

I. Introduction

Situation assessment or situation awareness (SA) is a key component of any decision-making process [1]. A good definition of SA is found at [2], [3]: *"Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and a projection of their status in the near future.*

In an information fusion process, situation assessment represents a high-level inference level to identify the likely situations given the observed events and obtained data. High-level information fusion studies theories and methods to effectively combine data from multiple sensors and related information to achieve more specific inferences that could be achieved by using a single, independent sensor.

Traditional approaches using observational data and *a priori* models are insufficient to deal with real-world complex problems. In this regard, an intelligent offshore oil industry environment is a very complex scenario. In this context, Occupational Health and Security (OHS) is a priority issue as it is an important factor to reduce the number of accidents and incidents. In the oil industry there exist standards to identify and record workplace accidents and incidents in order to provide guiding means on prevention efforts, indicating specific failures or reference, means of correction of conditions or circumstances that culminated in accident.

An information fusion model can help to intelligently predict undesirable events like accidents by taking into account the event location and time, historical data of past events and user's profile. A solution should involve a machine learning approach to learn from past anomaly events and to predict accidental events in time and space while also exploiting the contextual knowledge available.

Past knowledge (historical data) can be analyzed to find interesting patterns. For this, data mining is the most essential part of the knowledge discovery process which combines databases, artificial intelligence, machine learning and statistics techniques. The basic techniques for data mining include: decision tree induction, rule induction, artificial neural networks, clustering and association rules. Data mining can be applied to any domain where large databases are saved. Some applications are: failure prediction [4], biomedical applications [5], process and quality control [6].

Association rule learning is a popular and well-researched set of methods for discovering interesting relations between entities in a large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. Many algorithms for generating association rules were presented over time. Some well-known algorithms are Apriori [7], Eclat [8] and FP-Growth [9], but they only do half the job, since they are algorithms for mining frequent item sets. Another step needs to be done after to generate rules from frequent item sets found in a database.

In this work we describe an approach to deal with this problem involving high-level information fusion, ontologies and rule mining. Our proposed model provides the big picture about risk analysis for that employee at that place in that moment in a real world environment. Our contribution is to build a causality model for accidents investigation by means of a well-defined spatiotemporal constraints on offshore oil industry domain.

The paper is organized as follows. After providing an introduction to the OHS problem and the role of information fusion processes in building a risk picture, Section II briefly describes the state of the art and some application domains. Section III focuses on a knowledge retrieval model, its architecture, domain model and reasoning process. Section IV depicts the formalization of the mining information used by the context based reasoning process for threat detection and recognition. Subsequently, a case study involving data is presented in Section V. Finally, Section VI presents some conclusive remarks and outlines the current and future work been carried out in this area.

II. Foundations

Data fusion has been defined in [10] as "*a multi-level process dealing with the association, correlation, combination of data and information from single and multiple sources to achieve refined position, identify estimates and complete and timely assessments of situations, threats and their significance*. Data fusion (DF) and information fusion (IF) has been treated similarly in literature but when talking about data fusion it represents raw data, and when referring to information fusion, it implies a higher semantic level of fusion. The problem of information fusion has attracted significant attention in the artificial intelligence community, trying to innovate in the techniques used for combining the data and to refine state estimates and predictions.

Information Fusion can be classified depending on the level of abstraction [11]: low-level fusion, medium level fusion, high level fusion and multi-level fusion. In the process of low level fusion the raw data are directly provided as an input to the data fusion process. The medium level fusion is a feature level where features are fused to obtain other features that could be employed for other tasks. In the high level fusion a combination of symbolic representation is the entry of the fusion process. And in the multi-level fusion the entry comes from different levels of abstractions.

Others classifications are proposed: Dasarathys Functional Model [12] or JDL (Joint Directors of Laboratories) conceptual model proposed by the American Department of Defense [10]. The JDL classification model consists of five processing levels in the transformation of input signals to decision-ready knowledge. These levels are: level 0 or source pre-processing; level 1 or object refinement; level 2 or situation assessment; level 3 or impact assessment and level 4 or process refinement.

High-level fusion starts at level 2. Situation assessment (SA) aims to identify the likely situations given the observed events and obtained data. It establishes relationships between the objects. Relations (i.e., proximity, communication) are valued to determine the significance of the entities or objects in a specific environment. The aim of this level includes performing high-level inferences and identifying significant activities and events (patterns in general). The output is a set of high-level inferences. Situation assessment is an important part of the information fusion process because it is the purpose for the use of IF to synthesize the multitude of information, it provides an interface between the user and the automation, and (3) focuses data collection and management.

Intensive research has been done in past years focused on low-level information fusion, nowadays the focus is currently shifting towards high-level information fusion [13]. Compared to the increasingly mature field of low-level IF, theoretical and practical challenges posed by high-level IF are more difficult to handle. Some of the applications that involve high-level fusion are: Defense [14]–[18], Computer and Information Security [19], [20], Disaster Management [21]–[24], Fault Detection [25]–[27], Environment [28]–[30]. Also techniques for using contextual information in high-level information fusion architectures has been studied at [31].

In the context of oil and gas industry, is increasingly concerned with achieving and demonstrating good performance of occupational health and safety (OHS), through the control of its OHS risks, which is consistent with its policy and objectives. In the oil industry there exist standards to identify and record workplace accidents and incidents to provide guiding means on prevention efforts, indicating specific failures or reference, means of correction of conditions or circumstances that culminated in accident. So, events recognition is central to OHS, since the system can selectively start proper prediction services according to the user current situation and past knowledge taken from huge databases. In this sense, a fusion framework that combines data from multiples sources to achieve more specific inferences is needed. An information fusion system must satisfy the users functional needs and extend their sensory capabilities [32].

In fact, our proposal is inspired in the semantic strategy of Gomez et al. [31]. We actually propose a machine learning algorithm to learn from past anomaly events and to predict accidents events in time and space. It also use additional knowledge, like the contextual knowledge: user profile, event location and time, etc. Our proposed model provides the big picture about risk analysis for that employee at that place in that moment in a real world environment. Our contribution is to build a causality model for accidents investigation by means of a well-defined spatiotemporal constraints on offshore oil industry domain. Also, we use ontological constraints in the post-processing mining stage to prune resulting rules.

III. Context-Based Information fusion for Ambient Intelligence

In this section more details about the context-based information fusion model are provided. This is part of a previous work [33]. First, a detailed description of the proposed architecture, domain ontology and reasoning process described by means of inductive learning process.

A. Architecture

The architecture of our context-based fusion framework is depicted in Fig. 1. The context-aware system developed has a hierarchical architecture with the following layers: Services layer, Context Acquisition layer, Context Representation layer, Context Information Fusion layer and Infrastructure layer. The hierarchical architecture reflects the complex functionality of the system as shown in the following brief description of the functionality of particular layers:

- *Infrastructure Layer*: The lowest level of the location management architecture is the Sensor Layer which

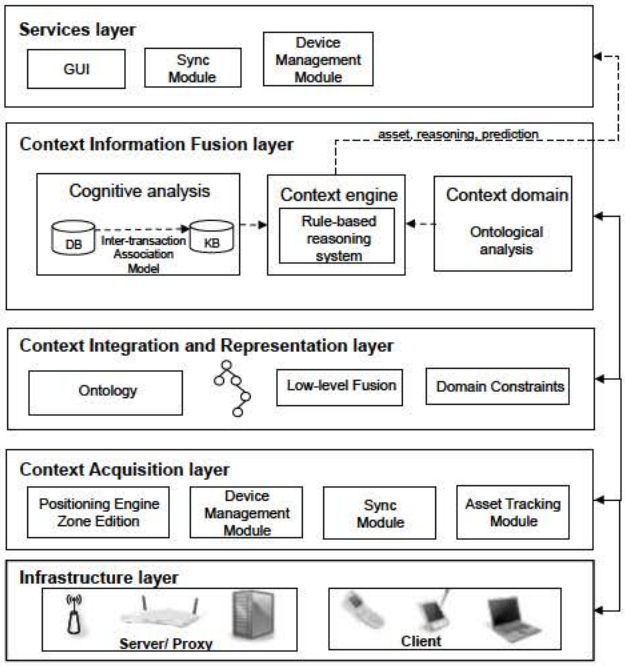

Fig. 1: Information fusion framework architecture.

represents the variety of physical and logical location sensor agents producing sensor-specific location information.

- *Context Acquisition*: The link between sensors (lowest layer) and the representation layer
- *Context Representation*: This is where the low-level information fusion occurs
- *Context Information Fusion Layer*: This layer takes sensor-specific location information and other contextual information related to the user and transforms it into a standard format. This is where the high-level information fusion occurs. It is here where reasoning about context and events of the past takes place. Extended description is given in next section.
- *Services Layer*: This layer interacts with the variety of users of the system (employees) and therefore needs to address several issues including access rights to location information (who can access the information and to what degree of accuracy), privacy of location information (how the location information can be used) and security of interactions between users and the system.

B. Ontology

Normally, an ontology represents a conceptualization of particular domains. In our case, we will use the ontology for representing the contextual information of the offshore oil industry environment.Ontologies are particularly suitable to project parts of the information describing and being used in our daily life onto a data structure usable by computers.

An ontology is defined as "an explicit specification of a conceptualization" [34]. An ontology created for a given domain includes a set of concepts as well as relationships connecting them within the domain. Collectively, the concepts and the relationships form a foundation for reasoning about the domain. A comprehensive, well-populated ontology with classes and relationships closely modeling a specific domain represents a vast compendium of knowledge in the domain.

Furthermore, if the concepts in the ontology are organized into hierarchies of higher-level categories, it should be possible to identify the category (or a few categories) that best classify the context of the user. Within the area of computing, the ontological concepts are frequently regarded as classes that are organized into hierarchies. The classes define the types of attributes, or properties common to individual objects within the class. Moreover, classes are interconnected by relationships, indicating their semantic interdependence (relationships are also regarded as attributes). We built a domain ontology for the Occupational Health and Security (OHS) (see Fig. 2) of oil and gas domain [35]. We also obtain the inferences that describe the dynamic side and finally we group the inferences sequentially to form tasks.

C. Reasoning

Standard ontology reasoning procedures can be performed within the ontologies to infer additional knowledge from the explicitly asserted facts. By using an inference engine, tasks such as classification or instance checking can be performed.

Risk prevention is a paradigmatic case of inductive reasoning. Inductive reasoning begins with observations that are specific and limited in scope, and proceeds to a generalized conclusion that is likely, but not certain, in light of accumulated evidence. You could say that inductive reasoning moves from the specific to the general. Much scientific research is carried out by the inductive method: gathering evidence, seeking patterns, and forming a hypothesis or theory to explain what is seen.

In our framework, inductive rules formally represent contextual, heuristic and common sense knowledge to accomplish high-level scene interpretation and low-level location refinement.

Once an employee enters the network, it immediately connects with a local proxy, which evaluates the position of the client device and assign a role to the employee. A pre-processing step begins then filtering the relevant features that are selected to participate in the process of knowledge discovery by type of employee (role). The association rules mining process starts with the selected configuration and the set of resulting rules can be analyzed. Later a post-processing step starts. It is an important component of KDD (knowledge-discovery in databases) consisting of many various procedures and methods for pruning and filtering the resulting rules.

The fusion engine implements an association rules model that combines dynamically feature selection based on the role of the user in order to find spatiotemporal patterns between different types of anomalies (or event sequence, ex. neglects, incidents, accidents) that match with the current location of the user.

Two categories of association mining are employed: intra-anomaly and inter-anomaly [36]. Intra-anomaly associations

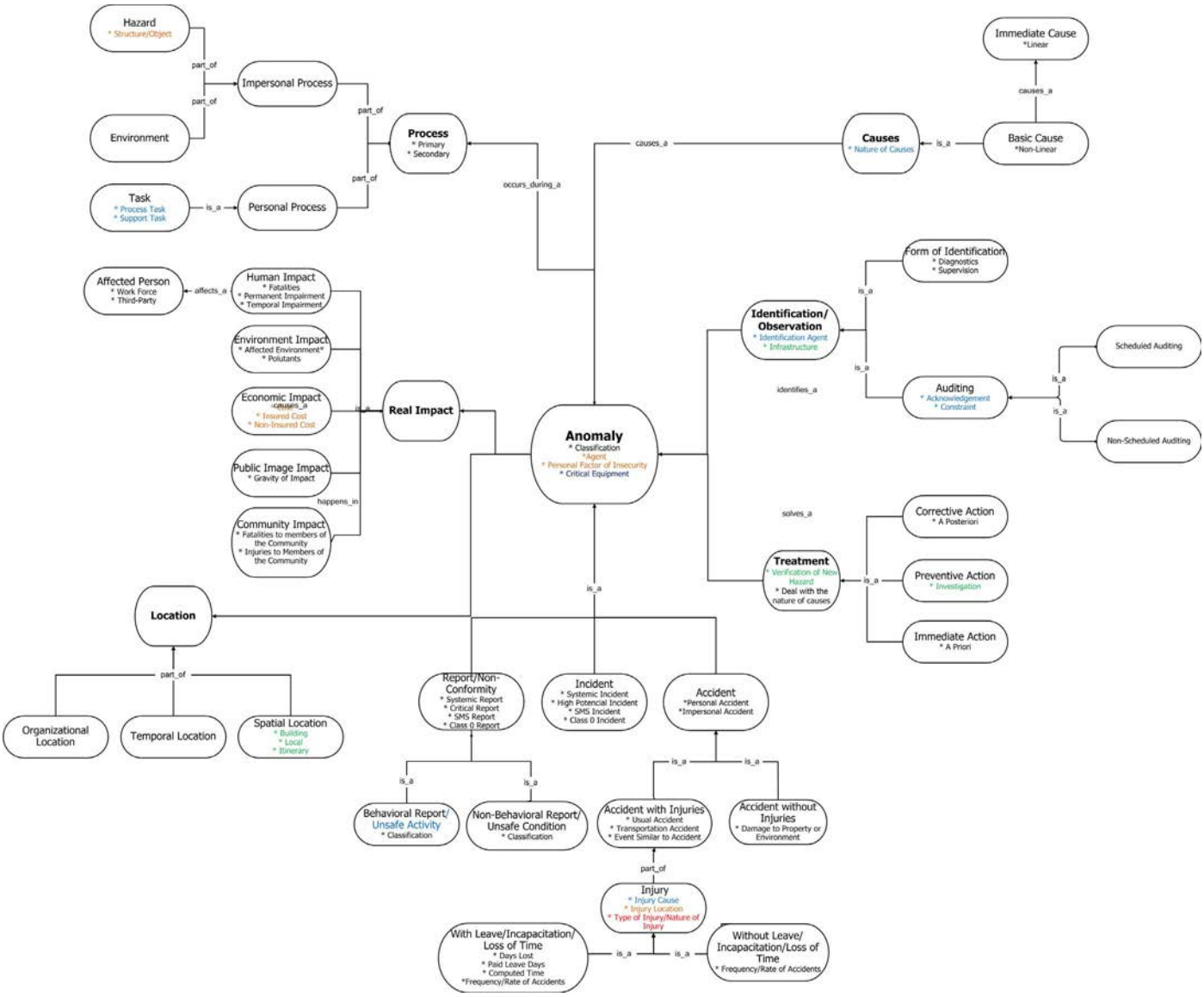

Fig. 2: Occupational Health and Security (OHS) ontology.

are the associating among items within the same type of anomaly, where the notion of the transaction could be events where the same user participate. However, inter-anomaly describes relationships among different transactions. That means between incidents, accidents and neglects. Further details are giving in the subsequent sections.

IV. Mining anomaly information

As already explained, the task of providing context-based information calls for the processing and extraction of information in the form of rules. One of the possible ways of obtaining those rules is to apply an association rule algorithm. In this work we employ Aprori and FP-Growth algorithms in parallel in order to mutually validate the results from each other.

As also explained in the above section, the fusion engine implements an association rules model that combines dynamically feature selection based on the role of the user in order to find spatiotemporal patterns between different types of anomalies (or event sequence, ex. neglects, incidents, accidents) that match with the current location of the user.

The dataset of anomalies, $\mathcal{S}$, is composed by anomaly instances,

$$\mathcal{S} := \{A_1, A_2, \ldots, A_n\}, n \in \mathbb{N}, \tag{1}$$

with the instances defined as

Definition 1 (Anomaly instance): An anomaly instance can be defined as the tuple,

$$A := \langle t, c, \mathcal{L}, \mathcal{O}, \mathcal{N}, \mathcal{F} \rangle, \tag{2}$$

that is composed by:

- t, a time instant that marks when the anomaly took place;
- $c \in \{\text{accident}, \text{incident}, \text{neglect}\}$, that sets the class of anomaly, and, therefore, its associated gravity;

- $\mathcal{L}$, a set of geo-location description attributes, which describe the geographical localization of the anomaly at different levels of accuracy;
- $\mathcal{O}$, a set of organizational location attributes that represent where in terms of organization structure the anomaly took place;
- $\mathcal{N}$, a set of descriptive nominal attributes that characterize the anomaly with a predefined values, and;
- $\mathcal{F}$, a set of free-text attributes that are used to complement or improve the descriptive power reachable with $\mathcal{N}$ attributes.

In order to make the rules produced interesting for the user the mining dataset, $\mathcal{S}$, must be preprocessed to meet the her/his needs. Using the above described problem ontology, the set of anomalies relevant for mining can be (i) filtered and (ii) its attributed selected.

For the first task we defined a function

$$\texttt{filter\_anomalies}\,(u, \mathcal{S}) \rightarrow \mathcal{S}', \mathcal{S}' \subseteq \mathcal{S}, \tag{3}$$

which determines the subset, $\mathcal{S}'$, of the anomalies dataset, $\mathcal{S}$, that are of interest for a given user, u. For the second task we created the function

$$\texttt{filter\_attributes}\,(u, \mathcal{S}') \rightarrow \mathcal{S}^*, \tag{4}$$

where $\forall A' =\in \mathcal{S}'$, $\exists A^* \in \mathcal{S}^*$ such that $t^* = t'$, $c^* = c'$, $\mathcal{L}^* \subseteq \mathcal{L}'$, $\mathcal{O}^* \subseteq \mathcal{O}'$, $\mathcal{N}^* \subseteq \mathcal{N}'$ and $\mathcal{F}^* \subseteq \mathcal{F}'$.

Relying on the $\mathcal{S}^*$ dataset customized to the user profile two classes of data mining operations can be carry out to extract knowledge rules. The first mines for rules regarding the relations of different attribute values in anomalies, and hence was called *intra-anomaly rule mining*. The other, more complex one, mines for relationships between anomalies, that take place in a same location —either geographical or organizational— and in similar dates. Because of that this operation was denominated *spatiotemporal* or *inter-anomaly rule mining*. In the subsequent sections we describe both mining processes.

A. Mining for intra-anomaly rules

In this case the data pre-processing before mining is pretty straightforward, as the interest is to discover relationships between the values of different attributes and the possible presence of probabilistic implication rules between them. In particular, each anomaly in $\mathcal{S}^*$ is treated as a transaction whose items are the non-null values of the corresponding $\mathcal{N}^*$. The descriptive attributes that take part of the mining process depend in the user profile. In order to model this we created a function the function

$$\mathcal{N}_{\text{sel}} = \text{select\_attributes}_{\text{intra}}\,(\mathcal{N}, u)\,, \tag{5}$$

which returns the subset of attributes, $\mathcal{N}_{\text{sel}} \subseteq \mathcal{N}$, that are of interest to a given user, u.

The results of applying the rule mining algorithms need to be post-processed to eliminate cyclic rules and to sort them according to an interestingness criterion. The outcome from this process should uncover relations between different values of the attributes. Some of those relationships might have a trivial

B. Mining for inter-anomaly rules

Mining spatiotemporal rules calls for a more complex preprocessing. As the most relevant anomalies are the accidents mining is centered around them. In this case, transactions will be constituted by anomalies that took place in the same location (deduced from the user profile) and with a given amount of time of precedence.

More formally, having the set of all accidents $\Lambda = \{A \in \mathcal{S}^* | A.c = \text{accident}\}$, for each element $\lambda \in \Lambda$, we construct the set of co-occurring anomalies, $\mathcal{C}\,(\lambda)$ as,

$$\begin{aligned}\mathcal{C}\,(\lambda) := &\{\kappa \in \mathcal{S}^* | \lambda.t - \kappa.t \leq \Delta t; \text{loc}\,(\lambda, u) = \text{loc}\,(\kappa, u)\} \\ &\cup \{\lambda\}\,,\end{aligned} \tag{6}$$

with $\text{loc}(\cdot)$, a function that for a given anomaly and user returns the value of the location attribute of interest for that user according to her/his role, and Δt, a time interval for maximum co-occurrence.

The set of co-occurring anomalies $\{\mathcal{C}(\lambda) | \forall \lambda \in \Lambda\}$ is used as transactions dataset for the mining algorithms. However, anomalies can not be used as-is, as it is necessary to express them in abstract form, in order to achieve sufficient generalization as to yield results that not are excessively particular or refined.

For this task, again depending on the user profile, a group of elements of each $\mathcal{N}^*$ is selected to create the abstract anomaly. This reduced set of attribute values are then used to construct the transactions. Therefore, as in the intra-anomaly case, we can construct a function

$$\mathcal{N}_{\text{sel}} = \text{select\_attributes}_{\text{inter}}\,(\mathcal{N}, u)\,, \tag{7}$$

that having given a user, u, returns the subset of attributes, $\mathcal{N}_{\text{sel}} \subseteq \mathcal{N}$, that are of interest to u. Relying on $\mathcal{N}_{\text{sel}}$, the abstracted anomaly A_{abstract} as the concatenation of the attribute/value pairs,

$$A_{\text{abstract}} = \bigoplus_{a \in \mathcal{N}_{\text{sel}}} a \oplus A.a, \tag{8}$$

where $\oplus$ is the concatenation operator. As this concatenated representation is inefficient from a computational point of view, they can be transformed into a reduced form by applying a hashing [37] or a compression [38] operator.

This process is better understood with an illustrative example. Fig. 3 puts forward such co-occurrence and abstraction process example. In this case, we have a simplified anomalies dataset $\mathcal{S}_{\text{sim}}$ where, having a given user, u, we also have the location attribute $l = \text{loc}\,(\lambda, u)$, $\forall \lambda \in \mathcal{S}_{\text{sim}}$ and the time of anomaly is denoted by t. For brevity reasons in the figure, the anomaly class is represented as $c = \{\texttt{A}, \texttt{I}, \texttt{N}\}$, for representing accidents, incidents and neglects, respectively, and the set of descriptive nominal attributes $\mathcal{N} = \{a_1, a_2, a_3\}$.

In this sample, there are three anomalies, which are marked with the A symbol. Assuming that $\mathcal{N}_{\text{sel}} = \{a_1, a_2\}$ (shaded in grey) and a $\Delta t = 2$, three transactions are created in the co-occurring anomalies dataset. This is because of that, for every accident, λ, $\lambda.c = \texttt{A}$ the anomalies that took place in the same location and with time in the interval $[\lambda.t - \Delta t, \lambda.t]$ are abstracted and added as a transaction to the mining dataset.

The resulting inter-anomaly mining dataset is composed by transactions that contain the abstracted version of the co-occurring anomalies for a given accident –or other class of anomaly of interest. As in the previous case, post-processing is necessary to filter out possible irrelevant and/or cyclic rules. For this a set of domain-principled filtering rules were proposed by the experts in order to define the most interesting consequent —accidents and incidents— and the preferred form of rules. As this is part is a sensitive element of the solution, involving trade decisions, we are not discussing it in detail.

V. Case Study

In this section we present a case study that was carried out with the intention of asserting from an experimental point of view the viability of the solution put forward in this work.

In order to create a controlled context for the tests it is required to (i) select a subset of the fused dataset, which contains all available anomalies, of such a size that can be directly handled by an expert and with such properties that guaranties the existence of rules (ii) create a custom user role that when applied selects a group of features for intra-anomaly mining and other for inter-anomaly mining.

In order to obtain the data set we applied a complex data query that filtered all accidents in a given time interval and their corresponding co-occurring anomalies. From that set, the accidents that did not have at least one more accident with the same abstracted co-occurring set were eliminated. These actions produced a set of about 2000 anomalies in which it was certain that there were latent rules relating some of them.

As the amount of anomalies in the dataset is of a manageable size the application of data visualization and inspection of software, along with the use of basic statistical tools allow to uncover at least some of the rules that are latent in the dataset. Therefore, two sets of expected rules were manually extracted with the purpose of verify that the mining algorithms were capable of discovering rules known to exist beforehand.

In all experiments the threshold parameters of the rule mining algorithm were set as: support, 0.2, and confidence, 0.8.

A. Intra-anomaly rule mining results

The application of FP-Growth in the intra-anomaly case yielded 71 frequent sets of items and 76 rules. The Apriori algorithm, in the other hand, generated 64 frequent item sets and 64 rules. An important analysis is to compared at what degree the frequent itemsets and rules generated by each approach overlaps the other. This can be posed as counting the number of itemsets and rules that have been generated by both methods. This comparison is presented in Table I. There it can be perceived that all itemsets and rules discovered by the Apriori method were also found by FP-Growth. This fact is a fundamental step to assert the validity of results.

Using the semi-automatic method explained above six rules extracted with the semi-automatic procedure. There rules were found to five of those rules were detected by Apriori, while only one by FP-Growth.

TABLE I: Similarity of the results produced by Apriori and FP-Growth in the intra- and inter-anomaly mining scenarios.

	Number shared	Shared Apriori results	Shared FP-Growth results
	— Intra-anomaly mining —		
Freq. item sets	64	100.0000%	90.1408%
Rules	64	100.0000%	84.2105%
	— Inter-anomaly mining —		
Freq. item sets	230	100.0000%	100.0000%
Rules	2670	100.0000%	100.0000%

TABLE II: Presence of the rules previously extracted by a semi-automatic procedure in the results of Apriori and FP-Growth algorithms.

	Apriori results	FP-Growth results
	— Intra-anomaly mining —	
Expected rules found	5	1
Coverage of expected rules	62.5000%	12.5000%
Per cent of true positives	7.8125%	$\approx$ 0.0000%
Per cent of non-expected rules found	92.1875%	100.0000%
	— Inter-anomaly mining —	
Expected rules found	6	6
Coverage of expected rules	100.0000%	100.0000%
Per cent of true positives	0.2247%	0.2247%
Per cent of non-expected rules found	99.7753%	99.7753%

B. Inter-anomaly rule mining results

After carrying out the process of abstraction and anomaly co-occurrence grouping a dataset with 1025 transactions was passed to the rule mining algorithms. Similarly, in that dataset, six rules were extracted by the semi-automatic method.

The results of both algorithms in this case are interesting. Table I show that Apriori and FP-Growth found the same number of frequent itemsets, 230, and of rules (before filtering), 2670. Again, this results validate the approach proposed.

When determining how many hand-drawn rules were actually found by the algorithms the results are also encouraging. All six rules were found by both algorithms as demonstrated in Table II.

VI. Final Remarks

In this work we have presented an information fusion framework for providing context-aware services related to risk prevention in offshore oil industry environment. The proposal put forward aims at providing context-based information related to accidents and their causes to users depending on their profiles and location.

Our approach relies on a domain ontology to capture the relevant concepts of the application and the semantics of the context in order to create a high-level fusion of information. Along with that we have introduced an innovative use of rule mining for provisioning knowledge for situation assessment and decision making regarding risk an accidents prevention. This form of rule mining is capable of an online high-level knowledge extraction that represents relations between different kinds of anomalies that have taken place at the user

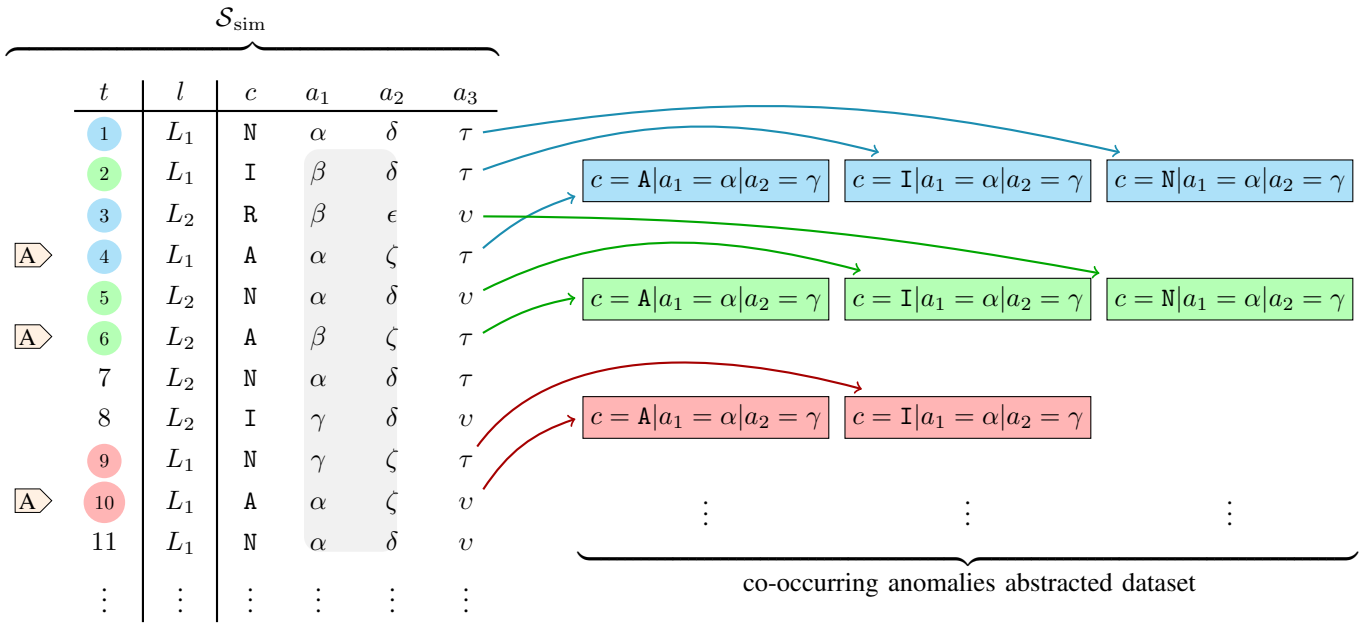

Fig. 3: Schematic representation of a co-occurrence and abstraction process example. The example is replies in a simplified anomalies dataset $\mathcal{S}_{\text{sim}}$ where, having a given user, u, the location attribute is $l = \text{loc}(\lambda, u)$, $\forall \lambda \in \mathcal{S}_{\text{sim}}$ and the time of anomaly is denoted by t. For brevity reasons, the anomaly class is represented as $c = \{\mathtt{A}, \mathtt{I}, \mathtt{N}\}$, for representing accidents, incidents and neglects, respectively, and the set of descriptive nominal attributes $\mathcal{N} = \{a_1, a_2, a_3\}$. In this sample, there are three anomalies, which are marked with the A symbol. Assuming that $\mathcal{N}_{\text{sel}} = \{a_1, a_2\}$ (shaded in grey in the schema) and a $\Delta t = 2$, three transactions are created in the co-occurring anomalies dataset. This means that, for every accident, λ, $\lambda.c = \mathtt{A}$ the anomalies that took place in the same location and with time in the interval $[\lambda.t - \Delta t, \lambda.t]$ are abstracted and added as a transaction to the mining dataset.

location and that the system has determined that had lead to an accident.

This feature has the potential of lowering at great length the development of accidents and incidents as it allows the users to directly act on the causes and conditions that have prompted such situations in the past. It empowers the users with the tools that help them to modify their routine and to avoid possible hazards or dangers.

This work is of particular relevance when taking into account the significant human, social, economical and environmental impact of accidents in this application context. From human and social points of view, the class of application described here is important as the remoteness and isolation of the installations render any assisting action more complicated and risky than usual. Similarly, oil industry is a heavily cost-minded industry, where accidents trend to have a important economical repercussions derived from the stop of production and the cost of the equipment and repair activities. Last, but certainly not least, the dramatic environmental impact of this industry has been sadly verified in the last years. Accidents, in the form of oil spills and fires are one of the main risks and one of the main dangers perceived by society regarding this industry. The environmental issue implies damages that are frequently impossible to assess in quantitative terms and whose footprint can potentially remains latent for future generations. It also has human, social and economic ramifications that fall in the above mentioned areas.

The solution presented here is currently deployed and in active use by a major oil extraction and processing industrial conglomerate of Brazil. It is currently used in the off-shore and inland oil extraction facilities as well as industrial and support locations.

The authors and collaborators are actively working in further improvements to the solution presented here. One important effort is directed towards the extension of the inter-anomaly relationship mining. The current model is being extended in order to be able to perform a multi-location mining. This work is being carried out in three main directions: (i) combine organizational and geographical localization; (ii) convert current location into a more abstract representation that could be generalizable across different installations, and; (iii) including geographical or topological neighbourhood information regarding the organizational or geographical location and relying on that extend the co-occurrence function definition. There is another important area or work that is focused on the integration of this framework in a multi-level information fusion framework. Similarly, the algorithms for knowledge discovery are currently being revised. Other machine learning paradigms like fuzzy inference, genetic programming or decision trees can be use to extract rules. They should be assessed and experimentally compared.

Acknowledgements

This work was partially funded by CNPq BJT Project 407851/2012-7 and CNPq PVE Project 314017/2013-5.

REFERENCES

[1] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, 1995.

[2] G. Bedny and D. Meister, "Theory of activity and situation awareness," *International Journal of Cognitive Ergonomics*, vol. 3, no. 1, pp. 63–72, 1999.

[3] K. Smith and P. Hancock, "Situation awareness is adaptive, externally directed consciousness," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 137–148, 1995.

[4] M. L. Borrajo, B. Baruque, E. Corchado, J. Bajo, and J. M. Corchado, "Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises," *International Journal of Neural Systems*, vol. 21, no. 04, pp. 277–296, 2011.

[5] J. F. De Paz, J. Bajo, V. F. López, and J. M. Corchado, "Biomedic organizations: An intelligent dynamic architecture for KDD," *Information Sciences*, vol. 224, pp. 49–61, 2013.

[6] M. Conti, R. D. Pietro, L. V. Mancini, and A. Mei, "Distributed data source verification in wireless sensor networks," *Information Fusion*, vol. 10, no. 4, pp. 342–353, 2009.

[7] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB '94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 487–499. [Online]. Available: http://dl.acm.org/citation.cfm?id=645920.672836

[8] M. J. Zaki, "Scalable algorithms for association mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, 2000.

[9] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *ACM SIGMOD Record*, vol. 29, no. 2. ACM, 2000, pp. 1–12.

[10] F. E. White, "Data fusion lexicon," DTIC Document, Tech. Rep., 1991.

[11] R. C. Luo, Y. C. Chou, and O. Chen, "Multisensor fusion and integration: algorithms, applications, and future research directions," in *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*. IEEE, 2007, pp. 1986–1991.

[12] B. V. Dasarathy, "Sensor fusion potential exploitation-innovative architectures and illustrative applications," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 24–38, 1997.

[13] E. Blasch, J. Llinas, D. Lambert, P. Valin, S. Das, C. Chong, M. Kokar, and E. Shahbazian, "High level information fusion developments, issues, and grand challenges: Fusion 2010 panel discussion," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[14] C.-Y. Chong, M. Liggins *et al.*, "Fusion technologies for drug interdiction," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI'94)*. IEEE, 1994, pp. 435–441.

[15] A. Gad and M. Farooq, "Data fusion architecture for maritime surveillance," in *Proceedings of the Fifth International Conference on Information Fusion (FUSION'02)*, vol. 1. IEEE, 2002, pp. 448–455.

[16] M. E. Liggins, A. Bramson *et al.*, "Off-board augmented fusion for improved target detection and track," in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 295–299.

[17] S. Ahlberg, P. Hörling, K. Johansson, K. Jöred, H. Kjellström, C. Mårtenson, G. Neider, J. Schubert, P. Svenson, P. Svensson *et al.*, "An information fusion demonstrator for tactical intelligence processing in network-based defense," *Information Fusion*, vol. 8, no. 1, pp. 84–107, 2007.

[18] T. Aldinger and J. Kao, "Data fusion and theater undersea warfare-an oceanographer's perspective," in *OCEANS'04. MTTS/IEEE TECHNO-OCEAN'04*, vol. 4. IEEE, 2004, pp. 2008–2012.

[19] I. Corona, G. Giacinto, C. Mazzariello, F. Roli, and C. Sansone, "Information fusion for computer security: State of the art and open issues," *Information Fusion*, vol. 10, no. 4, pp. 274–284, 2009.

[20] G. Giacinto, F. Roli, and C. Sansone, "Information fusion in computer security," *Information Fusion*, vol. 10, no. 4, pp. 272–273, 2009.

[21] E. G. Little and G. L. Rogova, "Ontology meta-model for building a situational picture of catastrophic events," in *8th International Conference on Information Fusion (FUSION'05)*, vol. 1. IEEE, 2005, pp. 1–8.

[22] J. Llinas, "Information fusion for natural and man-made disasters," in *Proceedings of the Fifth International Conference on Information Fusion (FUSION'02)*, vol. 1. IEEE, 2002, pp. 570–576.

[23] J. Llinas, M. Moskal, and T. McMahon, "Information fusion for nuclear, chemical, biological & radiological (NCBR) battle management support/disaster response management support," Center for MultiSource Information Fusion, School of Engineering and Applied Sciences, University of Buffalo, USA, Tech. Rep., 2002.

[24] J. Mattioli, N. Museux, M. Hemaissia, and C. Laudy, "A crisis response situation model," in *10th International Conference on Information Fusion (FUSION'07)*. IEEE, 2007, pp. 1–7.

[25] A. Bashi, "Fault detection for systems with multiple unknown modes and similar units," Ph.D. dissertation, University of New Orleans, 2010.

[26] A. Bashi, V. P. Jilkov, and X. R. Li, "Fault detection for systems with multiple unknown modes and similar units-part i," in *12th International Conference on Information Fusion (FUSION'09)*. IEEE, 2009, pp. 732–739.

[27] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using dempster-shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.

[28] U. Heiden, K. Segl, S. Roessner, and H. Kaufmann, "Ecological evaluation of urban biotope types using airborne hyperspectral hymap data," in *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*. IEEE, 2003, pp. 18–22.

[29] A. Khalil, M. K. Gill, and M. McKee, "New applications for information fusion and soil moisture forecasting," in *8th International Conference on Information Fusion (FUSION'05)*, vol. 2. IEEE, 2005, p. 7.

[30] L. Hubert-Moy, S. Corgne, G. Mercier, and B. Solaiman, "Land use and land cover change prediction with the theory of evidence: a case study in an intensive agricultural region of france," in *Proceedings of the Fifth International Conference on Information Fusion (FUSION'02)*, vol. 1. IEEE, 2002, pp. 114–121.

[31] J. Gómez-Romero, J. Garcia, M. Kandefer, J. Llinas, J. Molina, M. Patricio, M. Prentice, and S. Shapiro, "Strategies and techniques for use and exploitation of contextual information in high-level fusion architectures," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[32] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, and E. H. Ruspini, "Issues and challenges of knowledge representation and reasoning methods in situation assessment (level 2 fusion)," in *Defense and Security Symposium*. International Society for Optics and Photonics, 2006, pp. 623 510–623 510.

[33] N. Sanchez-Pi, L. Martí, J. M. Molina, and A. C. B. Garcia, "High-level information fusion for risk and accidents prevention in pervasive oil industry environments," in *J.M. Corchado et al. (Eds.): PAAMS 2014 Workshops, CCIS 430*. Springer International Publishing Switzerland, 2014, pp. 202–213.

[34] J. Gómez-Romero, M. A. Patricio, J. García, and J. M. Molina, "Ontological representation of context knowledge for visual data fusion," in *12th International Conference on Information Fusion (FUSION'09)*. IEEE, 2009, pp. 2136–2143.

[35] N. Sanchez-Pi, L. Martí, and A. C. Bicharra Garcia, "Text classification techniques in oil industry applications," in *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13*, ser. Advances in Intelligent Systems and Computing, A. Herrero, B. Baruque, F. Klett, A. Abraham, V. Snášel, A. C. Carvalho, P. García Bringas, I. Zelinka, H. Quintián, and E. Corchado, Eds., vol. 239. Springer International Publishing, 2014, pp. 211–220.

[36] C. Berberidis, L. Angelis, and I. Vlahavas, "Inter-transaction association rules mining for rare events prediction," in *Proc. 3rd Hellenic Conference on Artificial Intellligence*, 2004.

[37] A. L. Tharp, *File organization and processing.* Wiley, 1988.

[38] K. Sayood, *Introduction to Data Compression (2Nd Ed.)*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.